Annals of Clinical and Analytical Medicine
Original Research

Evaluating ChatGPT's responses on schizophrenia: Accuracy, quality, and stigmatization

Onur Gökçen[1], Ahmet Kağan Misci[2]
[1] Department of Psychiatry
[2] Medical Student, Faculty of Medicine, Kütahya Health Science University, Kütahya, Turkey

Abstract

Aim: ChatGPT, developed by OpenAI, is an AI-powered language model widely used for online information retrieval. This study aims to assess the accuracy, sufficiency, and stigmatization of ChatGPT's responses on schizophrenia as evaluated by psychiatrists using validated assessment scales.

Material and Methods: Common patient queries were identified based on frequently asked questions in clinical practice, relevant healthcare platforms, and previous patient experiences. The questions were formulated with input from ten psychiatry specialists and then posed to ChatGPT in a single session. Psychiatrists evaluated the responses using the Global Quality Scale (GQS), the DISCERN Scale, and a stigma-related question.

Results: Among 57 psychiatrists (12 residents, 45 specialists), GQS scores indicated "good to excellent" quality, with higher ratings for treatment-related questions. Responses were generally not perceived as stigmatizing. The DISCERN total score averaged 50, classifying responses as "good," though the lowest ratings were given for source credibility and date information.

Discussion: To the best of our knowledge, this is the first study assessing ChatGPT's reliability as an information source on schizophrenia. It provides a valuable reference for future research on the accuracy, reliability, and ethical considerations of evolving large language models in psychiatry.

Keywords

Schizophrenia, Artificial Intelligence, ChatGPT, Patient Information

DOI: 10.4328/ACAM.22665 Received: 2025-03-20 Accepted: 2025-05-05 Published Online: 2025-05-28 Printed: 2025-06-01 Ann Clin Anal Med 2025;16(6):450-454
Corresponding Author: Onur Gökçen, Department of Psychiatry, Faculty of Medicine, Kütahya Health Science University, Kütahya, Turkey.
E-mail: onurgokcen29@gmail.com P: +90 507 201 52 21
Corresponding Author ORCID ID: https://orcid.org/0000-0003-2058-9855
Other Authors ORCID ID: Ahmet Kağan Misci, https://orcid.org/0009-0000-7794-0322
This study was approved by the Ethics Committee of Kütahya Health Sciences University (Date: 2024-10-24, No: 2024/12-28)

Introduction

Schizophrenia is a complex and chronic neuropsychiatric disorder that affects approximately 1% of the global population. It is characterized by a variety of symptoms, including positive symptoms (such as hallucinations and delusions), negative symptoms (such as anhedonia and social withdrawal), and cognitive impairments (such as deficits in working memory and executive function) [1]. Schizophrenia typically emerges in early adulthood and leads to significant impairments in social and occupational functioning, resulting in a reduction in life expectancy by approximately 15 years compared to the general population [1, 2]. Schizophrenia patients seek information about their condition from various sources, including the internet, their doctors, other healthcare professionals, and their social networks. Lower educational attainment has been associated with increased misconceptions regarding the etiology of schizophrenia, often leading to supernatural attributions [3]. Access to accurate information plays a critical role in helping patients and their families understand the causes, treatment options, and coping strategies for schizophrenia [4]. Psychoeducation is an essential component of treatment that improves care quality by reducing stigma [5].

Despite increased awareness about schizophrenia, surveys show that stereotypical beliefs about patients, such as tendencies toward violence, maladaptive behaviors, and an inability to maintain employment, remain widespread. The media often portrays schizophrenia with negative stereotypes, influencing public perception. This association between schizophrenia and violence further reinforces stigma [6, 7]. As patients internalize this stigma, their social isolation may increase, leading to higher stress levels and worsening of symptoms [8]. This underscores the importance of reliable and accessible information to ensure a proper understanding of the illness [4].

The internet is used as a source of mental health information by more than 10% of the general population, and by over 20% of individuals who have experienced mental health issues [9]. Approximately 24% of users consider the internet to be one of the top three most reliable sources of information. Due to the accessibility and anonymity it provides, individuals at risk of stigma are more likely to turn to the internet for health information. Additionally, many individuals first identify their symptoms online before seeking medical assistance [10].

Recently, AI-powered language models have become an important tool for obtaining information online. One such model, ChatGPT, is a large language model developed by OpenAI that can generate human-like responses on various topics. Trained on vast datasets, ChatGPT typically produces accurate and informative answers using advanced machine learning techniques [11]. However, as AI chatbots inherently lack the ability to independently verify the accuracy of the information they provide, they cannot guarantee the most up-to-date or comprehensive information. Nevertheless, they are increasingly being used by patients to gather information about illnesses and treatments. The potential for generating inaccurate or misleading information remains a significant concern [12, 13].

Studies examining the use of AI language models as reliable sources of information have addressed topics such as vaccine and statin hesitancy, clinical psychiatry knowledge, and spinal surgery [11, 14, 15]. However, there is no research evaluating whether AI models can serve as a source of information on schizophrenia or examining the stigmatizing nature of ChatGPT's responses. This study aims to evaluate the information provided by ChatGPT regarding schizophrenia, as assessed by psychiatrists in terms of accuracy, sufficiency, and stigmatization.

Material and Methods

In this study, the most frequently asked questions by patients with schizophrenia to obtain information were identified through the examination of websites created by schizophrenia associations, medical associations, and healthcare institutions that provide services, as well as by evaluating the patients' previous experiences. Subsequently, the questions were refined based on the opinions of ten experienced psychiatrists. The questions are presented in Table 1.

Following this, these questions were individually directed to ChatGPT, developed by OpenAI, in a single session. For the evaluation of the responses, the Global Quality Scale (GQS) and the DISCERN scale were used, as in similar studies [16, 17].

DISCERN (Criteria for Quality of Consumer Health Information): DISCERN is a scale developed to assist patients and healthcare providers in assessing the quality of health information. This scale consists of 15 questions in total, each scored on a scale from 1 (low quality) to 5 (high quality) [18]. The first eight questions assess the reliability of the information, while the following seven focus on treatment options. The final, 16th question is used to determine the overall quality of the information by considering all previous evaluations. The total score for the first 15 questions is classified as excellent (63-75), good (51-62), moderate (39-50), inadequate (28-38), and very inadequate (15-27) [19].

Global Quality Scale (GQS)

The GQS is another scale used to evaluate the quality and usefulness of health information from the patient's perspective. This five-question scale assesses factors such as the integrity, clarity, and practical utility of the information (Table 1). The total quality score is calculated by summing the scores of each section. Content with a total GQS score of 3 or below is considered of low to medium quality, while content with scores above 3 is classified as good to excellent quality [20]. When used in conjunction with DISCERN, GQS provides a more holistic evaluation in terms of the comprehensibility and patient guidance of the information [17].

A form was created to allow separate evaluations of each question and its corresponding answer from ChatGPT using the GQS, and a collective assessment of the total responses using the Discern scale. Additionally, an extra evaluation question, created by the researchers, was added to the form to examine the responses to the eighth, ninth, and tenth questions in terms of "stigmatization."

"Evaluate the responses to questions 8, 9, and 10 in terms of stigma:

Does not stigmatize

Partially stigmatizes

Stigmatizes

This form was distributed to psychiatry specialists via Google

Forms. The responses provided by a total of 57 psychiatry specialists were analyzed in this study. Among them, 12 were residents, while the remaining 45 were specialists in the field.

Ethical Approval

This study was approved by the Ethics Committee of Kütahya Health Sciences University (Date: 2024-10-24, No: 2024/12-28).

Results

Answers to questions about schizophrenia were provided by ChatGPT. The quality of each answer was evaluated by psychiatrists using the Global Quality Scale (GQS) (table 1). The average Global Quality Scale (GQS) scores for the responses to the questions were found to be in the "good to excellent quality" range, with scores above 3. The responses to the sixth and seventh questions, which were related to treatment, had higher average GQS scores. A Kruskal-Wallis test was conducted to determine whether there was a statistically significant difference between the GQS scores for the responses. The analysis revealed a significant difference (p = 0.014). Epsilon-

Table 1. Global quality scale

Score	Global Score Description
1	Poor quality, poor flow of the site, most information missing, not at all useful for patients
2	Generally poor quality and poor flow, some information listed but many important topics missing, of very limited use to patients
3	Moderate quality, suboptimal flow, some important information is adequately discussed but others poorly discussed, somewhat useful for patients
4	Good quality and generally good flow, most of the relevant information is listed, but some topics not covered, useful for patient
5	Excellent quality and excellent flow, very useful for patients

Table 2. Scores Assigned to ChatGPT's Answers Based on the Global Quality Scale

Questions	Mean	SD	p	H
1. What is schizophrenia, and what are its symptoms?	3.47	0.87	0.014*	20.665
2. How is schizophrenia diagnosed?	3.73	0.72		
3. What causes schizophrenia? What are the reasons behind it?	3.67	0.80		
4. Is there a treatment for schizophrenia, and what are the treatment options?	3.80	0.87		
5. Are schizophrenia medications dangerous, and do they cause dependency?	3.49	1.6		
6. In what situations might hospitalization be necessary?	4.4	0.60		
7. How long does schizophrenia treatment last? Is it necessary to take medication for life?	4.2	0.78		
8. Are schizophrenia and people with schizophrenia dangerous?	3.89	0.78		
9. Can one be friends with someone who has schizophrenia? Can a romantic relationship or marriage be possible?	3.53	1.6		
10. Does schizophrenia prevent people from working? Can people with schizophrenia pursue any profession they wish	3.82	0.83		
Evaluate the answers to questions 8, 9 and 10 in terms of stigmatization	1.98	0.91		

Data expressed in mean ± SD, * p<0.05, H: Kruskal-Wallis test statistic

Table 3. Evaluation of ChatGPT's Responses Based on the DISCERN Scale

DISCERN (Quality Criteria for Consumer Health Information)	Mean	SD	Min	Max
1. Are the aims clear?	4.1	0.71	3.00	5.00
2. Does it achieve its aims?	3.75	0.74	2.00	5.00
3. Is it relevant?	4.13	0.99	2.00	5.00
4. Is it clear what sources of information were used to compile the answers?	1.93	1.11	1.00	5.00
5. Is it clear when the information used or reported in the publication was produced?	1.78	0.99	1.00	5.00
6. Is it balanced and unbiased?	3.77	0.95	1.00	5.00
7. Does it provide details of additional sources of support and information?	1.77	1.20	1.00	5.00
8. Does it refer to areas of uncertainty?	3.15	0.1	1.00	5.00
9. Does it describe how each treatment works?	2.71	0.92	1.00	5.00
10. Does it describe the benefits of each treatment?	3.26	0.89	2.00	5.00
11. Does it describe the risks of each treatment?	2.95	0.90	1.00	5.00
12. Does it describe what would happen if no treatment is used?	3.20	1.9	1.00	5.00
13. Does it describe how the treatment choices affect overall quality of life?	3.20	1.12	1.00	5.00
14. Is it clear that there may be more than one possible treatment choice?	3.80	1.14	1.00	5.00
15. Does it provide support for shared decision-making?	3.28	1.16	1.00	5.00
Based on the answers to all of the above quetionsquestions, rate the overall quality of the publication as a source of information about treatment choices	3.33	0.92	2.00	5.00
Discern total score	50	9.77	35	80

Data expressed in mean ± SD, Min= Minimum, Max=Maximum

Squared (ε^2) was calculated to be 0.0265, indicating a small effect size. A post-hoc Dunn's test showed that a significant difference was only found between the responses to the first and sixth questions ($p < 0.001$).

Participants were instructed to evaluate the responses to the eighth, ninth, and tenth questions in terms of stigma. It was observed that, on average, participants did not find the responses to be stigmatizing (Table 2).

In the evaluation using the Discern scale, the average total score for the participants was found to be 50. According to the standards of this scale, this indicates that the responses were rated as "good." The first eight questions assess the reliability of the information, while the following seven focus on treatment options. Among the first eight Discern items assessing the reliability of the information, the lowest scores were given to items four, five, and seven. These questions were related to the sharing of the sources and dates of the information (Table 2). Among the last seven questions, which focus on treatment options, the lowest score was given to the response to the eleventh question. This question inquired whether the risks of each treatment were clearly defined.

Discussion

In this study, the responses provided by ChatGPT regarding schizophrenia were evaluated by 57 psychiatrists using the Global Quality Scale (GQS) and DISCERN scales. The evaluation revealed that, overall, the responses provided by ChatGPT were sufficient. Thus, this research will contribute meaningfully to the ongoing discussions about the use of large language models in psychiatry and expand potential areas for future studies.

The use of artificial intelligence (AI) in psychiatry has attracted significant interest from researchers. In recent years, large language models (LLMs) have paved the way for many innovative studies in the field of psychiatry. Previous studies in this context have demonstrated the success of LLMs in generating new and analyzable data points from original data using data augmentation techniques on complex and imbalanced text-based data found on social media platforms [21].

Moreover, through a specially developed framework, the capacity of LLMs to analyze text-based data and detect mental health conditions, as well as provide personalized and sensitive commentary, has been assessed, showing significant achievements in this area [22]. Additionally, the potential applications of LLMs in psychodynamic analysis and psychoanalytic formulations have been explored using written expressions from patients with different psychiatric diagnoses [23].

Overall, the number of studies focusing on the use of LLMs in healthcare is increasing, and the results emerging from these studies appear to be consistent and promising. GPT models are standing out and are widely preferred in these studies. Moreover, each new study lays the groundwork for future research and contributes to the advancement of the field.

The use of large language models as a source of information in psychiatry is also of significant importance. This study is the first to evaluate the quality of ChatGPT's responses regarding schizophrenia; however, several studies have evaluated ChatGPT as a source of information in various medical fields.

The average Global Quality Scale (GQS) scores of the responses to the questions were found to be in the "good to excellent quality" range. The highest score was given to the response to the question 'In what situations might hospitalization be necessary?', while the lowest score was given to the response to the question 'What is schizophrenia, and what are its symptoms?'. The difference in scores between these two responses was found to be statistically significant. The question about the necessity of hospitalization may be seen as a more straightforward topic with fewer required details, whereas the definition and symptoms of schizophrenia can be described in a much broader and more detailed manner. Furthermore, certain pieces of information that may be considered minor details could have been deemed important by psychiatrists, and their absence in the responses might have led to those responses being rated as 'insufficient'. Indeed, the literature emphasizes that GPT's accuracy can be highly variable, especially on specific topics such as rare diseases, and that this variability may be related to the limitations of online sources and GPT's difficulty in selecting appropriate references [11, 24]. Moreover, previous studies examining the capabilities and limitations of GPT have shown that, although the model can produce non-conspiratorial responses to questions about vaccines and statins, it generates text based on linguistic associations, and the way questions are phrased can influence the quality of the responses [11, 24]. Indeed, compared to a more open-ended question such as 'What is schizophrenia?', a question like 'In what situations might hospitalization be necessary?', which allows for a clear, itemized response, may lead to answers that are more straightforward and easier to evaluate.

Similarly, in different medical specialties, the success rate of GPT has shown variability. For example, in a study conducted in the field of otolaryngology, it was noted that GPT achieved vastly different success rates across sub-specialties, performing better in more popular sub-specialties while showing lower success rates in less popular ones [25].

In our study, generally, more common and frequently encountered questions were selected. Therefore, our study's results do not provide sufficient confidence regarding the quality of responses to more specific and rare questions. Furthermore, upon examining the DISCERN scale results, it was observed that the sources of information were not sufficiently clear and detailed in the responses provided by ChatGPT. Like other AI-assisted language models, ChatGPT is continuously evolving, and there may have been advancements in the provision and sharing of sources since the time this study was conducted. The GPT-3.5 model used in our study is free and is actively preferred by approximately 180 million users. While it is anticipated that more advanced and consistent responses will be obtained with future models, it is not guaranteed that later versions such as GPT-4 will always produce better results.

Limitation

This study examined ChatGPT's responses on schizophrenia using two evaluation scales. It also explored whether ChatGPT can be considered a reliable information source. However, the study does have some limitations. First and foremost, ChatGPT, like other AI-assisted language models, is continuously

developing. Therefore, the model's performance and the quality of its responses may vary over time. Moreover, the responses provided by ChatGPT may change depending on the updates to online sources, meaning that responses obtained after the time this study was conducted may differ. Another important limitation is that the study was conducted exclusively in Turkish. While ChatGPT is capable of multilingual output, its performance can vary significantly across languages due to differences in training data volume, linguistic structure, and cultural context. These language-specific variations may affect the quality, completeness, and nuance of responses. Therefore, the results of this study may not be generalizable to responses generated in other languages, particularly English, which tends to be more prominently represented in ChatGPT's training data.

Conclusion

Despite these limitations, this study is the first to evaluate the potential of ChatGPT as a reliable source of information on schizophrenia. The findings highlight both the possible benefits and the points of caution when using AI-assisted language models in medical information. In this context, the study provides an important reference point that can lay the groundwork for future research. Future studies could provide valuable insights into how these technologies can be used more effectively in healthcare by evaluating various AI models more comprehensively in terms of medical accuracy, source usage, and reliability.

Scientific Responsibility Statement

The authors declare that they are responsible for the article's scientific content including study design, data collection, analysis and interpretation, writing, some of the main line, or all of the preparation and scientific review of the contents and approval of the final version of the article.

Animal and Human Rights Statement

All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or compareable ethical standards.

Funding: *None*

Conflict of Interest

The authors declare that there is no conflict of interest.

References

1. McCutcheon RA, Marques TR, Howes OD. Schizophrenia—an overview. JAMA Psychiatry. 2020;77(2):201-10.

2. Liu X, Wang D, Fan R, Wang R, Xiang H, Yang X, et al. Life expectancy and potential years of life lost for schizophrenia in western China. Psychiatry Res. 2022;308:1-5.

3. Suryani S, Ningsih EW, Nuraeni A. Knowledge, perception, and burden of family in treating patients with schizophrenia who experience relapse. Belitung Nurs J. 2019;5(4):162-8.

4. Chiu Y-H, Kao M-Y, Goh KK, Lu C-Y, Lu M-L. Renaming schizophrenia and stigma reduction: a cross-sectional study of nursing students in Taiwan. Int. J Environ Res Public Health. 2022;19(6):3563.

5. Laçiner K, Şenol Y. How easy is it to read websites about schizophrenia? Cureus. 2023;15(11):1-7.

6. Owen PR. Portrayals of schizophrenia by entertainment media: a content analysis of contemporary movies. Psychiatr Serv. 2012;63(7):655-9.

7. Li J, Zhang M-m, Zhao L, Li W-q, Mu J-l, Zhang Z-h. Evaluation of attitudes and knowledge toward mental disorders in a sample of the Chinese population using a web-based approach. BMC Psychiatry. 2018;18:1-8.

8. Lysaker PH, Davis LW, Warman DM, Strasburger A, Beattie N. Stigma, social function and symptoms in schizophrenia and schizoaffective disorder: Associations across 6 months. Psychiatry Res. 2007;149(1-3):89-95.

9. Powell J, Clarke A. Internet information-seeking in mental health: population survey. Br J Psychiatry. 2006;189(3):273-7.

10. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med. 2005;61(8):1821-7.

11. Torun C, Sarmis A, Aytekin O. Is ChatGPT an accurate and reliable source of information for patients with vaccine and statin hesitancy? Medeniyet Med J. 2024;39(1):1.

12. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183(6):589-96.

13. Semrl N, Feigl S, Taumberger N, Bracic T, Fluhr H, Blockeel C, et al. AI language models in human reproduction research: exploring ChatGPT's potential to assist academic writing. Hum Reprod Open. 2023;38(12):2281-8.

14. Luykx JJ, Gerritse F, Habets PC, Vinkers CH. The performance of ChatGPT in generating answers to clinical questions in psychiatry: a two□layer assessment. World J Psychiatry. 2023;22(3):479.

15. Stroop A, Stroop T, Zawy Alsofy S, Nakamura M, Möllmann F, Greiner C, et al. Large language models: are artificial intelligence-based chatbots a reliable source of patient information for spinal surgery? Eur Spine J. 2024;33(11):4135-43.

16. Kurt Demirsoy K, Buyuk SK, Bicer T. How reliable is the artificial intelligence product large language model ChatGPT in orthodontics? Angle Orthod. 2024;94(6):602-7.

17. Tirumala AKG, Mishra S, Trivedi N, Shivakumar D, Singh A, Shariff S. A cross-sectional study to assess response generated by ChatGPT and ChatSonic to patient queries about Epilepsy. Telemat Inform Rep 2024;13: (1-5).

18. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999;53(2):105-11.

19. Yılmaz R, Karpuz S, Yılmaz H, Solak İ. Osteoporoz ile ilgili türkçe web sitelerinin bilgi içeriği, okunabilirlik, güvenilirlik ve kalitesinin değerlendirilmesi [Evaluation of information content, readability, reliability and quality of turkish websites related to osteoporosis]. Turk J Osteoporos. 2023;29(2):109-116.

20. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Van Zanten SV. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol. 2007;102(9):2070-7.

21. Liyanage C, Garg M, Mago V, Sohn S, editors. Augmenting reddit posts to determine wellness dimensions impacting mental health. Proc Conf Assoc Comput Linguist Meet. 2023;1:306–12.

22. Mazumdar H, Chakraborty C, Sathvik M, Panigrahi PK. GPTFX: A novel GPT-3 based framework for mental health detection and explanations. IEEE J Biomed Health Inform. 2023;1:1-8.

23. Hwang G, Lee DY, Seol S, Jung J, Choi Y, Her ES, et al. Assessing the potential of ChatGPT for psychodynamic formulations in psychiatry: An exploratory study. Psychiatry Res. 2024;331:1-7.

24. Valentini M, Szkandera J, Smolle MA, Scheipl S, Leithner A, Andreou D. Artificial intelligence large language model ChatGPT: is it a trustworthy and reliable source of information for sarcoma patients? Front Public Health. 2024;12:1-6.

25. Hoch CC, Wollenberg B, Lüers J-C, Knoedler S, Knoedler L, Frank K, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. Eur Arch Otorhinolaryngol. 2023;280(9):4271-8.

How to cite this article:

Onur Gökçen, Ahmet Kağan Misci. *Evaluating ChatGPT's responses on schizophrenia: Accuracy, quality, and stigmatization. Ann Clin Anal Med 2025;16(6):450-454*

This study was approved by the Ethics Committee of Kütahya Health Sciences University (Date: 2024-10-24, No: 2024/12-28)